ICoGESF (2025)
https://doi.org/xxxx/xxxx/xxxx

The 1st International Conference on Green Engineering for Sustainable Future

Air Quality Forecasting in Makassar Using the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) Method

Dimas Dwi Pramono[1], Pebri Andika Putra[2], M Galih Rahmatulloh[3], and Muh Farhan Baidlowi[4]

1Department of Informatics Engineering, Faculty of Engineering, State University of Surabaya,

Ketintang Campus, Ketintang Street, Surabaya 60231, East Java, Indonesia

2Department of Informatics Engineering, Faculty of Engineering, State University of Surabaya,

Ketintang Campus, Ketintang Street, Surabaya 60231, East Java, Indonesia

3Department of Informatics Engineering, Faculty of Engineering, State University of Surabaya,

Ketintang Campus, Ketintang Street, Surabaya 60231, East Java, Indonesia

4Department of Informatics Engineering, Faculty of Engineering, State University of Surabaya,

Ketintang Campus, Ketintang Street, Surabaya 60231, East Java, Indonesia

Abstract. Today's society really pays attention to air quality because the impact of exposure to pollutants in the air is starting to be felt. PM 2.5 pollutants are very dangerous because their small size can penetrate the alveoli of human lungs. The value calculation of the Air Quality Index (AQI) is important to prepare mitigation and defensive measures to reduce the negative impact of air quality and as a basis for future policymaking. Several method comparisons have been carried out by researchers to predict AQI. However, researchers have not studied much regarding the use of meteorological factors in the form of average air temperature (°C), average air humidity (percent), and average wind speed (m/s) in forecasting AQI values, even though meteorological factors have a significant link, according to previous researchers. This research forecasts AQI using the ARIMAX method, which includes meteorological factors as exogenous variables, using daily AQI PM 2.5 data in Central Jakarta. The best modeling of the data is ARIMA (1,1,1) without X and ARIMAX (1,1,1). Based on the calculation of AIC, BIC, RMSE, and MAPE values, ARIMAX (1,1,1) modeling produces better forecasting, so it can be concluded that forecasting involving meteorological factors can make forecasting more precise. Predicting AQI using ARIMAX with upcoming meteorological factors is beneficial, as precise prediction results can assist in policy- making to prevent the adverse impacts of air quality on public health. In future research, other meteorological factors could be studied and combined with other modeling besides ARIMA

1 Introduction

Air quality is a crucial environmental indicator, especially in urban areas with rapid industrialization and transportation activities. Makassar City has experienced increased pollution levels, particularly PM2.5 particles, which can penetrate deep into the respiratory tract and cause health problems such as cardiovascular and respiratory diseases. Predictive studies suggest that the Air Pollution Standard Index (ISPU) in Makassar will rise significantly by 2045, highlighting the need for accurate forecasting models.

This research is conducted to:

- Predict daily PM2.5 concentrations using the ARIMA model.
- Analyze the impact of meteorological variables on PM2.5 predictions using the ARIMAX model.

1.1.1 Problem Description

The main problem in air quality management in Makassar is the lack of accurate predictions regarding changes in pollutant levels over time. Historical data cannot be utilized optimally in producing predictions that can be used for strategic decision making. This results in preventive actions being taken that are not optimal.

1.1.2 Purpose Problem

The following are the objectives of our research:

- Designing an air quality forecasting model in Makassar using the ARIMAX method.
- Making predictions of PM 2.5 air quality values based on historical data.
- Making predictions of PM 2.5 air quality values based on historical data and external variables.

1.1.3 Previous Solutions

Several previous studies have been conducted on air quality prediction in Indonesia and other countries. Common methods usually applied are linear regression, simple time series models, and there is also the use of machine learning such as Random Forest and Support Vector Machine (SVM). However, most of these studies only use historical data without considering external variables that can affect PM 2.5 pollutant levels, such as water intensity, humidity, weather, and traffic activity. In addition, the application of the ARIMAX model as a forecasting method that combines historical data and external variables is still limited, especially in Makassar.

1.1.4 Lack of Solutions

Currently, the limitation of the solution lies in the integration of external variables with historical data is still lacking in the air quality prediction process. Most of the applied models only focus on past patterns without accommodating external factors that can influence PM 2.5 pollutants.

1.1.5 Problem Formula

The formulation of the problem in this study is:

- How to make predictions of PM 2.5 air quality values based on historical data.
- How to make predictions of PM 2.5 air quality values based on historical data and external variables.
- How to design an air quality forecasting model in Makassar using the ARIMA method.

1.1.6 Solution Offered

This study offers the development of a PM 2.5 air quality forecasting model in Makassar using the ARIMAX (AutoRegressive Integrated Moving Average with eXogenous variables) method. The designed model combines historical air quality data with external variables, such as weather data, traffic density, and industrial activity. Therefore, with this approach, the resulting predictions can be more accurate and reliable as a basis for strategic decision making in air quality management.

1.1.7 Expectations Solution

The solutions offered through research are as follows:

- Providing an accurate and responsive PM 2.5 air quality prediction model to external variables.
- Supporting the Makassar regional government in taking preventive and responsive steps to reduce the impact of air pollution on the community.
- Becoming a reference in developing an air quality prediction system in a city.

2. Theoretical Framework

2.1 Air Quality and PM2.5

Air quality is commonly evaluated using PM2.5, PM10, NO_2, SO_2, and CO levels. PM2.5 is particularly dangerous due to its microscopic size and deep lung penetration (Pope & Dockery, 2006).

2.2 Augmented Dickey-Fuller (ADF) Test

The ADF test checks time series stationarity. If non-stationary, data must undergo differencing for ARIMA modeling (Widarjono, 2018).

2.3 ARIMA Model

ARIMA (p,d,q) incorporates autoregression (AR), differencing (I), and moving average (MA). It is suitable for non-seasonal, univariate time series data (Box & Jenkins, 1976).

2.4 ARIMAX Model

ARIMAX extends ARIMA by including exogenous variables that influence the dependent variable, such as temperature and humidity (Hyndman, 2022).

2.5 Forecast Evaluation Metrics

- MSE: Mean squared prediction error.
  RMSE: Root of MSE, in original units.
  MAPE: Mean absolute percentage error, useful for comparing model performance.

3. Methodology

3.1 Data Collection

PM2.5 Data: Retrieved from AQICN via public API for Makassar.
Meteorological Data: Includes daily average temperature (TAVG), humidity (RH_AVG), and wind speed (FF_AVG) from BMKG.
Period: February 7 to May 24, 2025.
Missing Data: Handled using hybrid mean interpolation.

3.2 Analysis Steps

1. Exploratory Data Analysis: Data merging and trend identification via plots.
2. ARIMA Modeling: Conducted with ADF test and ACF/PACF plots to identify (p,d,q).
3. ARIMAX Modeling: Developed using R with external regressors (TAVG, RH_AVG, FF_AVG).
4. Model Evaluation: Models assessed with MAE, RMSE, and MAPE.
5. Visualization: Actual vs. predicted PM2.5 plotted for interpretability.

4. Results and Discussion

4.1 Data Analysis

Data was cleaned, structured as time series, and processed using R. Exploratory analysis revealed daily variation patterns in PM2.5 and meteorological variables.

TANGGAL	TAVG	RH_AVG	FF_AVG	PM2.5
07-02-2025	25.1	92	3	6.10625
08-02-2025	26	91	5	6.3575
09-02-2025	24	96	3	7.89125
10-02-2025	27.2	89	6	8.13875
11-02-2025	25.7	94	3	9.55
12-02-2025	26	91	2	10.4225
13-02-2025	27.2	87	2	9.7
14-02-2025	27.7	85	2	8.7875
15-02-2025	27.6	84	2	9.105
16-02-2025	27.4	83	1	10.00125
17-02-2025	27.2	86	1	12.4
18-02-2025	27.4	85	2	10.46
19-02-2025	27.4	87	2	8.04875
20-02-2025	26.7	89	2	5.31625
21-02-2025	26.9	85	2	4.57875
22-02-2025	27.4	82	2	7.8416875
23-02-2025	27.7	78	2	6.86271875
24-02-2025	25.5	88	2	5.88375
25-02-2025	27	83	1	5.225
26-02-2025	26.9	85	2	5.87
27-02-2025	26.1	88	1	4.7475

4.2 ARIMAX Model Construction

The ARIMAX model incorporated TAVG, RH_AVG, and FF_AVG. Stationarity was achieved using differencing. Model diagnostics confirmed the residuals met the assumptions.

Hasil Koefisien:	ar1	TAVG	RH_AVG	FF_AVG
	0.5844	0.4426	-0.0397	-0.4153

4.3 Forecast Evaluation

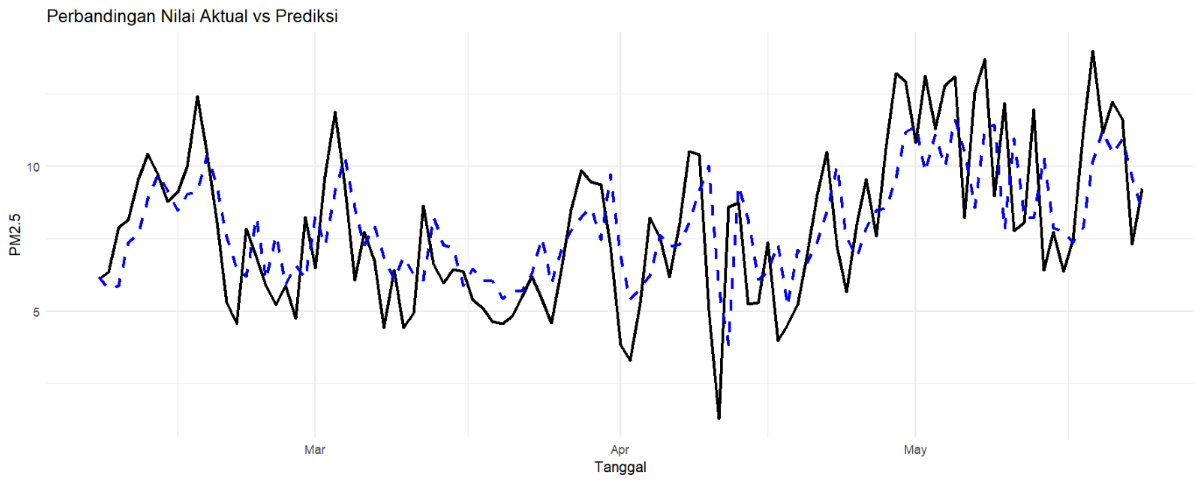
MSE	RMSE	MAPE
4.117047691	2.0290509	25.38226537

MSE: 4.117
RMSE: 2.029
MAPE: 25.38%

The relatively low RMSE and moderate MAPE indicate that the model has good predictive accuracy. The inclusion of exogenous variables improved performance compared to standalone ARIMA.

4.4 Visualization

Prediction plots show the model closely follows actual PM2.5 trends, with only minor deviations at extreme points.



5. Conclusion

This study successfully modeled PM2.5 air quality in Makassar using ARIMA and ARIMAX methods. The ARIMAX model provided better forecasting results, demonstrated by RMSE of 2.029 and MAPE of 25.38%. It confirms that integrating meteorological factors enhances prediction accuracy, making ARIMAX suitable for real-world air quality forecasting and policy planning.

References

Akbar, M. (2016). BAB II Tinjauan Pustaka. Repository ITK.
Basuki, A., & Prawoto, H. (2016). BAB III Metode Penelitian. Repository Untidar.
Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. Holden-Day.
Digilib Unila. (2017). Uji kestasioneran mengenai data inflasi. Universitas Lampung.
Hyndman, R. J. (2022). Discussion on ARIMAX model and exogenous variables in air quality forecasting.
Jusri. (2023). Makassar city air pollution forecast in 2045. Hasanuddin University.
Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi). (2024). Forecasting the air quality index by utilizing several meteorological factors using ARIMAX method.
Keramidas, T., Tchung-Ming, S., Weitzel, M., & Van Dingenen, R. (2023). Analisis kualitas udara dan dampak PM2.5 terhadap kesehatan masyarakat. Jurnal Ilmu Lingkungan, 21(2), 381–386.
Kuncoro, M. (2017). Metode riset untuk bisnis dan ekonomi. Erlangga.
Moshammer, H., & Wallner, P. (2011). Air quality as respiratory health indicator—A critical review. International Journal of Occupational Medicine and Environmental Health, 24(3), 241–248. https://doi.org/10.2478/s13382-011-0028-9
Pope, C. A., III, & Dockery, D. W. (2006). Health effects of fine particulate air pollution: Lines that connect. Journal of the Air & Waste Management Association, 56(6), 709–742. https://doi.org/10.1080/10473289.2006.10464485
Ramdhan, A., & Syamsuddin, S. (2020). Urban growth and industrial development in Makassar City: Impacts on environmental sustainability. Journal of Environmental Management and Tourism, 11(4), 789–796.
Repository UIN Suska. (2022). BAB III. Universitas Islam Negeri Sultan Syarif Kasim Riau.
Utama, D. A. (2019). Indeks standar pencemar udara polutan karbon monoksida di Terminal Malengkeri Kota Makassar. Jurnal UNHAS.
Wahid, S., & Setyawan, H. (2020). Peramalan data deret waktu menggunakan model ARIMA. Indonesian Journal of Applied Statistics, 3(2), 136–147. https://jurnal.uns.ac.id/ijas/article/view/26298/19416
Widarjono, A. (2018). Ekonometrika: Teori dan aplikasi untuk ekonomi dan bisnis (Edisi ke-3). UPP STIM YKPN.
World Health Organization. (2021). WHO global air quality guidelines: Particulate matter (PM2.5 and PM10), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. World Health Organization.